

Selection of the Informative Feature System for Crops Classification Using Hyperspectral Data

S. M. Borzov^{1*} and O. I. Potaturkin^{1**}

¹*Institute of Automation and Electrometry, Siberian Branch, Russian Academy of Sciences,
Novosibirsk, 630090 Russia*

Received May 15, 2020; revised May 22, 2020; accepted June 8, 2020

Abstract—The methods based on processing video data have shown their efficiency in many areas of agriculture and forestry. Nevertheless, they are not enough accurate for classification of barely discernible objects and types of plants, which can be provided only using hyperspectral sensors. However, such devices by now were expensive and difficult in operation and were used mostly on satellites and manned aircraft. In recent years, the technologies for creating smaller and lighter sensors have been proposed; these technologies are based on the choice of restricted number of spectral intervals and their position at the design stage. They may be used for scientific or commercial goals under field conditions and, in addition, installed on unmanned aerial vehicles. Exemplified by a 220-channel hyperspectral image, this paper investigates the capability of significant reduction in the amount of registered data due to choice of position and width of restricted number of most informative spectral channels in solving the classification problem of agricultural plants. It is shown that the applied method for generating feature systems has a considerable advantage against the regular decimation and is close in its efficiency to the methods based on the analysis of principal components, but has a significantly less amount of needed computations.

DOI: 10.3103/S8756699020040032

Keywords: Earth's remote sensing, hyperspectral images, classification of agricultural plants, choice of informative feature.

1. INTRODUCTION

Today, the methods and means of Earth's remote sensing (ERS) are being actively developed based on capabilities of application of hyperspectral (HS) technologies. The differential characteristic of images registered in such recording is the large number of channels in the visible, near infrared, and shortwave infrared ranges [1, 2]. Taking into account uniqueness of absorption spectrum of different substances, the registered information potentially allows determining not only the type of observed objects, but also their state (for instance, the content of individual pigments and moisture in the vegetation cover). This significantly increases the amount of data and, therefore, the difficulty of their transmission, storage, and processing. At the same time, not all registered information is identically fruitful. Many studies demonstrated the property of inter-channel redundancy of HS images [3]. Moreover, spectral signatures may be subject to variations in dependence on illumination and atmospheric conditions [4]. This leads to necessity of development of the technique and technology for determining the most informative spectral channels.

In addition to that, in recent years there were several attempts of using advantages of HS technologies in creating mobile monitoring systems intended to work under the field conditions. It is difficult due to the fact that the traditional HS cameras are based on application of dispersion element, scanner, and matrix receiver [5]. They scan the object by the slot diaphragm and use the first order of diffraction to decompose the images into a spectrum. At each step of scanning, a single section of HS image is

*E-mail: borzov@iae.nsk.su

**E-mail: potaturkin@iae.nsk.su

registered; therefore, the time of generation of the complete image with such camera counts to several seconds or even minutes. Note that such approach is convenient in solving the ERS problems, where scanning is organized by moving the vehicle along the orbit. These devices are not appropriate to most practical applications, where the camera or the objects of monitoring have arbitrary spatial motions.

The key novelty of recent years has been the refusal of necessary application of dispersion element for dividing the input radiation into a set of bands with different wavelengths. This has been realized using the idea of depositing narrow spectral filters directly on the surface of high-quality and simultaneously mass-produced CMOP sensor on the level of semiconductor plate [6–10]. This technology has been called the filter-on-chip. All this has led to appearance of hyperspectral filter organized in a tiled manner, that is, divided into equal rectangular sections each of which is responsible for its own narrow spectral range. For instance, the two-megapixel CMOSIS CMV2000 sensor includes 32 tile sections with the resolution of 256×256 pixels each. In total, the camera forms the image with the resolution of 256×256 pixels with the rate of approximately 30 frames/s under day light and up to 340 frames/s under better conditions applied, for instance, in machine vision. The camera is sensitive in the visual and near infrared ranges (600–1000 nm) with the width of channels of 12 nm. The next step in developing these technologies is the mosaic filter [11]. In this case the mosaic HS sensor is divided into the groups of pixels each of which has a certain spectral filter and the resulting HS image is formed from the corresponding pixels of different groups.

The above mentioned technologies make HS cameras comparatively reasonable for multiple applications: they are sufficiently compact, light, and suitable for using as monitoring systems of various intention. Such systems may be adjusted for specific application by determining the necessary number and widths of spectral bands and the resolution of image in each spectrum range. The choice of spectral intervals, of their number and position is usually performed at the system design stage and does not vary during operation [6–9]; however, there are well-known attempts of creating the filters with adjustable characteristics [10]. During studies of the imaging filtering hyperspectrometer, it was shown that, despite a nonideal transmission function of a filter, the error of generation of spectral distribution is in general moderate and is not larger than 13% [12].

In works [13–15], in processing of large-sized HS images, the efficiency of several methods for spectral and spectral–spatial classification of plant types was studied with different number of features. As a result, it was shown that it is possible to use a small number of features (15–20 for natural zone and 10–15 for urban area) generated by using the principal component method and its various modifications [16, 17]. It allowed reducing the computation time by two orders of magnitude without considerable decrease in the classification accuracy. However, for generation of the mentioned features we still need to conduct recording, transmission, and storage of data of hundreds of spectral channels, because the resulting features are computed by linear combination of all original ones.

The current work is aimed at experimental study of the possibility of significant reduction in the amount of registered data due to selection of a restricted number of most informative spectral channels in solving the problem of classification of agricultural plants. The studies are carried out on an example of processing of data from a 220-channel aviation hyperspectral image obtained in the AVIRIS program in the Indian Pine test site [18].

2. CHOICE OF INFORMATIVE SUBSYSTEM OF FEATURES

The task consists in the following: to form the most informative system consisting of n features from N features by analyzing the training set (set of pixels with classes known due to ground observations) ($n < N$). This requires checking $N!/(N-n)!n!$ alternative combinations, which may be done by simple exhaustive search for a small N . However, the growth in the number of features leads to exponential increase in the computation time, which is of crucial importance in designing systems with adjustable filters.

The simplest way to avoid necessity of exhaustive search is the technique of sequential truncation of features (the Del algorithm) [19]. When this algorithm is applied, we remove one of features in turn and determine the information content of resulting systems of $N - 1$ features. As the information content criterion we use the accuracy of classification of the training set, which is computed as a fraction of correctly classified pixels in percent. Thus, we determine the most informative system of $N - 1$ features and repeat the procedure once again to achieve n features.

The similar approach is used in the algorithm of sequential addition of features (the Ad algorithm) [20]. It is different from the previous one only by the fact that the order checking the subsystems of features begins not from N -dimensional spaces, but from one-dimensional spaces. Firstly, all N features are tested for their information content. To this end, the training set is classified into each feature individually and the informative subsystem includes the feature providing the largest accuracy of classification. After that, one of remaining features is added to it in turn and the most informative subsystem of two features is selected. The process is repeated until the system with required number of features is obtained.

Both described algorithms give the optimal solution at each step, but this does not provide the global optimum. To introduce a random element, many studies apply the methods of stochastic and evolution optimization (for instance, the genetic algorithms (GA) and the particle swarm optimization algorithms (PSO) [21, 22]), which act using random choice, combination, and variation in sought parameters using the mechanisms similar to the natural selection.

In the proposed work, to weaken the effect of errors at the first steps of procedure, with the Ad algorithm we collect some number of informative features and then eliminate some of them using the Del algorithm. The repetition of these procedures (the AdDel algorithm) is continued until the system with required number of features is obtained.

Note that the resulting set of spectral channels depends on the applied method of classification at the stage of evaluation of the information content of the formed systems of features at algorithm's intermediate stages. In this work we carry out such evaluation using the maximum likelihood (ML) [23] and the support vector machine (SVM) [24, 25] methods. The ML method is based on estimating the second-order statistic (covariation matrix) and requires large sizes of training sets (TSs) for HS data. For correct estimation the size of the training set must be ten times larger than the number of spectral channels. In the situation when the TS size is limited, it is possible to obtain degenerate (singular) covariation matrices. The SVM method does not apply the second-order statistic and can be used for a TS of limited size. However, the SVM also has its own disadvantages in its application for our goals. Firstly, its classification accuracy is somewhat lower for large number of features and TS size than that of ML. Secondly, it is based on the iteration procedure of testing multiple alternative decisions for adjusting the parameters of the dividing hypersurface and, as a consequence, requires plenty of CPU time and other computational resources. Thus, the method for estimating the information content of systems of features is extremely important and it should be chosen following from the parameters of the processed data and available computational resources.

3. SELECTION OF WIDTH OF SPECTRAL CHANNELS

After selection of spectral channels, it is purposeful to determine their width. The presence of 200-channel HS image allows adjusting the width of spectral channels providing the maximum separability of the training set. To this end, we average the data of each selected channel together with several neighboring ones and evaluate the informative content of the resulting system. We repeat the procedure for different number of averaged channels and determine the number at which the maximum classification accuracy of the training set data is achieved. The selection of the width is cyclically repeated for each channel at fixed widths of other channels chosen previously and is terminated when they stop varying.

It is established that the proposed algorithm leads to different resulting combinations at different choice of original widths of selected channels.

To avoid the trap of local optimum, we apply random prescription of original widths for selected channels. We test several random variants and choose the one providing the best result in using the procedure of the width selection.

In the application of chosen methods for generating systems of features, there is no absolute assurance in the fact that they are capable of finding the global optimum; however, the current work is aimed at demonstrating the proposed approach, but not at finding the technique of choosing the optimal system of features. For the techniques of choosing the optimal system of features, see, e.g., works [26, 27].

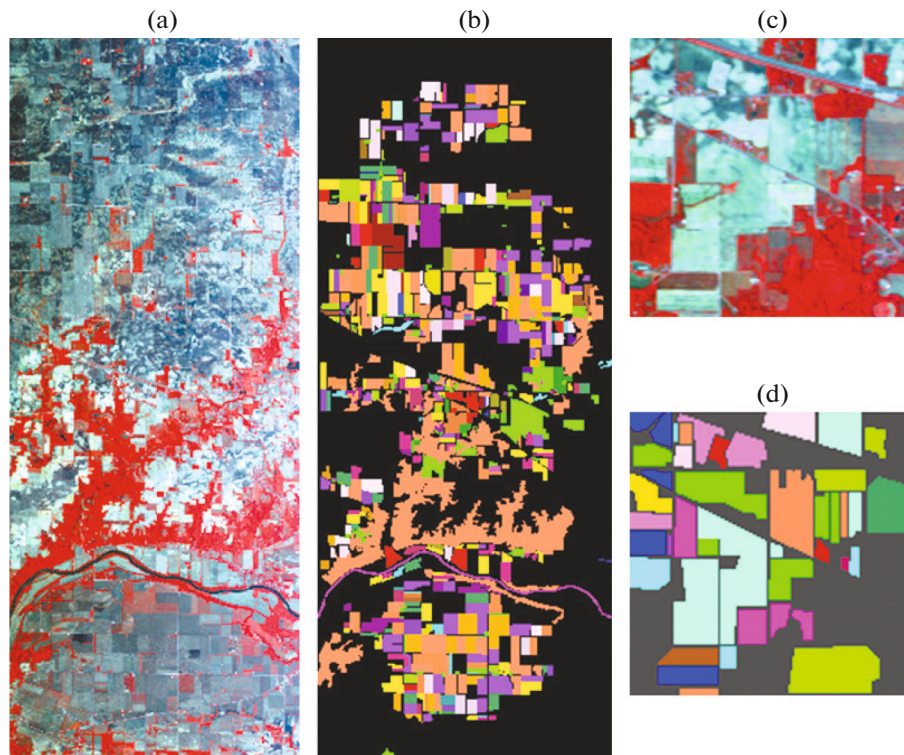


Fig. 1. Initial data of ERS: (a) RGB composite of a large-sized HS image (channels 40 : 20 : 10), (b) reference map of image classes, (c) RGB composite of a fragment, (d) reference map of fragment classes.

4. EXPERIMENTAL RESULTS

The experimental studies of the efficiency of data classification based on reduced subsystems of features were performed with application of a large-sized HS image (Fig. 1a). The image size is 614×2677 pixels, its resolution is 20 m/pixel, and the number of channels in the range $0.4\text{--}2.5 \mu\text{m}$ is 220; using the results of ground (subsattellite) observations, the image is divided into 58 classes (see the reference map of classes in Fig. 1b). They include the agricultural plants (including 15 corn classes and 18 soy classes different in soil cultivation methods), highway, railway, forest, and residential facilities.

To generate the informative subsystems of features, we used a fragment of image with the size of 145×145 pixels containing 16 classes (Figs. 1c and 1d), 14 of which are the different plant classes, including 3 corn classes and 3 soy classes. We introduced 25% pixels of each class present in the fragment into the training set.

According to the AdDel algorithm, we generated consecutive subsystem of features in the cycle. As features we used the brightnesses of spectral channels of the HS image. The system of features generated at the previous cycle was consecutively supplemented with the previously unselected channels, and the classification of TSs of the mentioned fragment and the estimation of its accuracy were carried out for each combination of features. Following from this, we determined the subsystem which provides the maximum classification accuracy with a current number of features. Here, three cycles of increasing system of features were alternated with a single cycle of its reduction in order to obtain the required number of features.

In Table 1 we provide the numbers of 10 channels selected as a result of execution of this procedure at evaluating the information content of the systems of features with application of the ML and SVM classification methods. Figure 2 shows the generated subsystems with 10 and 20 features on the background of spectrum of a soy subclass for illustrative purposes. We can see that the sets of channels for the ML and SVM methods are not identical, but are close in their character. As expected, the most informative spectral range for separating plant types is the spectral range approximately from 0.45 to $0.76 \mu\text{m}$ [28] (channels nos. 23–37). In all cases the density of selected channels is maximal in this range.

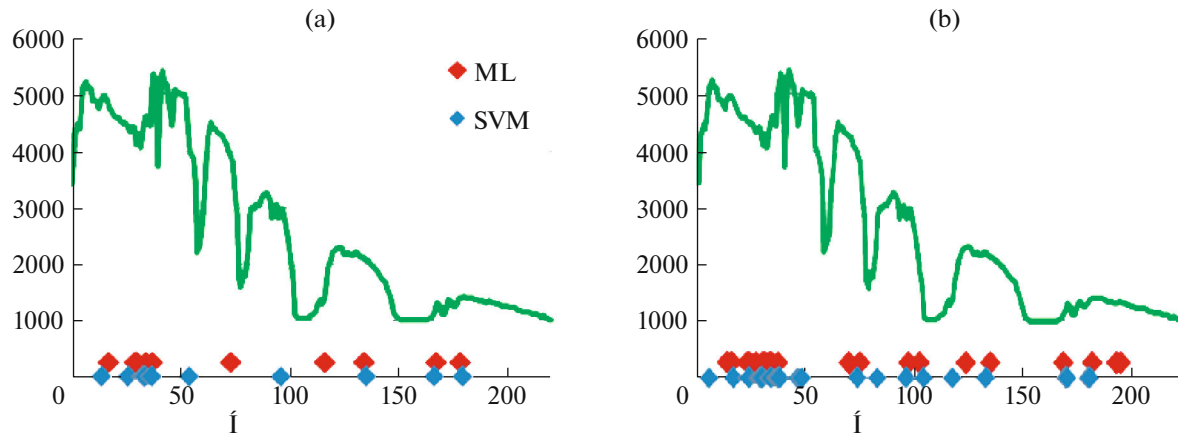


Fig. 2. Spectral channels chosen by AdDel algorithm from 220 initial channels at application of ML and SVM for estimating information content of systems of features: (a) 10 channels and (b) 20 channels.

This range includes the reflection maximum of the foliage in the visible range ($\sim 0.54 \mu\text{m}$) and a sharp growth in the reflection coefficient (approximately from $0.70 \mu\text{m}$) in the near infrared range of spectrum. These effects are caused by a relatively low absorption of chlorophyll (the absorption maximum is at bands of 0.45 and $0.65 \mu\text{m}$) containing in a healthy leave.

Each plant type is characterized by its own content of chlorophyll (and other pigments), and this leads to appearance of certain peculiarities in the absorption and reflection spectrum in a given region of wavelengths. Thus, if the plant is damaged and is under stress, the generation of chlorophyll reduces dramatically, which leads to a decrease in absorption in these bands. This fact allows determining the areas of vegetation cover affected by the negative impact [29].

After the selection of the prescribed number of most informative features, the spectral width providing the maximally efficient separability of TSs was chosen for each of them (by averaging the data of the adjacent channels). The width was chosen in two stages. At the first stage the width of the selected channel was determined so that it gave the maximum classification accuracy of the training set while the widths of all remaining channels were prescribed in a certain manner. At the second stage the width selection procedure was cyclically repeated for each channel with the widths of other channels chosen previously and terminated when they stop changing. We have already mentioned that this procedure was executed for several variants of prescribed initial widths. As the resulting set, we used the set of widths that provides the maximum efficiency of classification of the training set.

In Table 2 we give the widths of spectral channels obtained by applying the proposed procedure with the minimum original values prescribed for the case of selection of 10 channels (we estimated the informative content with the ML classification method). From now on, the numerical value of the width determines the number of adjacent channels unified with the channel selected by the AdDel algorithm: the value 0 means that only the data of the chosen channel are taken into account, the value n means that the data of $2n + 1$ channels are averaged (n with smaller and n with larger index).

Table 1. Choice of channels by AdDel algorithm

Channels in order of choice										
ML	74	133	14	37	24	179	33	96	166	34
SVM	134	73	29	37	34	178	167	17	116	30
Channels ordered by indices										
ML	14	24	33	34	37	74	96	133	166	179
SVM	17	29	30	34	37	73	116	134	167	178

Table 2. Width selection of ten channels chosen by AdDel algorithm

Index of channel	74	133	14	37	24	179	33	96	166	34
Initial width	0	0	0	0	0	0	0	0	0	0
Width (1 stage)	5	9	3	10	3	0	0	1	0	0
Width (2 stage)	8	2	4	6	3	0	6	10	5	1

Table 3. Width selection of ten channels chosen by AdDel algorithm at different initial widths

Initial combination of widths of 10 chosen channels	Accuracy of separation of TSs at initial widths	Resulting combination of widths	Accuracy of separation of TSs at resulting widths
0, 0, 0, 0, 0, 0, 0, 0, 0, 0	84.83	8, 2, 4, 6, 3, 0, 6, 10, 5, 1	88.02
1, 1, 1, 1, 1, 1, 1, 1, 1, 1	85.15	8, 4, 4, 7, 2, 0, 8, 7, 7, 1	88.09
10, 2, 4, 6, 2, 2, 7, 8, 1, 1	87.54	9, 2, 6, 6, 2, 3, 8, 7, 3, 1	88.21
9, 4, 6, 8, 9, 2, 9, 7, 0, 1	86.64	9, 2, 6, 6, 2, 3, 8, 7, 3, 1	88.21
5, 9, 3, 10, 1, 1, 1, 1, 1, 1	86.25	9, 2, 6, 10, 1, 5, 6, 10, 2, 1	87.23
5, 9, 3, 10, 3, 0, 0, 1, 0, 0	85.93	7, 10, 0, 4, 2, 0, 2, 1, 4, 1	86.88
8, 1, 10, 4, 1, 1, 2, 1, 7, 6	84.75	9, 6, 2, 6, 2, 0, 6, 10, 8, 1	87.94
1, 2, 3, 4, 5, 6, 7, 8, 9, 10	83.10	10, 9, 2, 7, 1, 9, 4, 6, 1, 1	87.23
5, 5, 5, 5, 5, 5, 5, 5, 5, 5	82.99	9, 10, 6, 7, 3, 0, 6, 10, 4, 1	87.94
3, 4, 4, 3, 2, 1, 0, 0, 1, 2	84.72	9, 2, 6, 6, 2, 3, 8, 7, 3, 1	88.21
3, 3, 3, 3, 3, 3, 3, 3, 3, 3	83.89	9, 7, 10, 9, 1, 4, 5, 5, 3, 1	87.23
2, 2, 2, 2, 2, 2, 2, 2, 3, 2	84.91	2, 10, 4, 1, 1, 4, 8, 5, 2, 1	87.31

Table 4. Classification of image fragment by ML method (by TSs) with application of subsystems of features

Number of channels	Decimation	PCA	MNF	AdDel	AdDel + choice of width
10 channels	74.7 (77.7)	80.0 (82.8)	86.3 (88.1)	79.7 (84.2)	84.4 (88.2)
20 channels	79.0 (85.9)	85.9 (90.6)	88.8 (92.2)	84.7 (91.9)	87.7 (93.1)
40 channels	81.3 (95.8)	85.2 (95.9)	87.8 (96.5)	82.9 (97.2)	86.4 (97.7)

In Table 3 we present the resulting combinations of widths and the informative content for different choices of initial widths of the corresponding channels.

Note that even rather close initial combinations often give different results (in the considered example fifth and sixth rows). At the same time, the third, the fourth, and the tenth combinations led to the same resulting set of widths which provides the maximum accuracy of separation of the training set. This set was used in subsequent studies.

The systems of 10, 20, and 40 features were formed (by the AdDel algorithm and by the AdDel algorithm with selection of channel widths) and further used in classification of an image fragment. The resulting accuracies are given in Table 4. For comparison's sake we also present here the data obtained in generating the system of features by the regular decimation, by the principal component analysis (PCA), and by the principal component analysis with the minimum noise fraction (MNF) [30]. By the table, it

Table 5. Classification of full image by ML method with application of subsystems of features (merging of classes of corn and soy)

Number of channels	Decimation	PCA	MNF	AdDel	AdDel + choice of width
10 channels	57.8	63.2	67.6	62.7	64.9
20 channels	63.1	69.6	72.4	65.0	70.1
40 channels	68.7	71.6	74.3	68.9	72.4
100 channels	74.3	74.8	76.5	–	–

follows that the applied method for generating subsystems of attributes has a considerable advantage against the regular decimation and is near in its efficiency to the methods based on the principal component analysis (in particular, to MNF). However, as we have already mentioned, its undisputed advantage is the possibility of decreasing requirements to the registering apparatus, bandwidth capacity of data transmission channels, and required computational performance.

Note that the maximum classification accuracy of the test set of a fragment is achieved when 20 features are applied. As the number of features is further increased, the separation accuracy of the training set increases, but this effect is obviously associated with overtraining and has no practical value, since the classification accuracy of the test set begins decreasing, which is explained by insufficient representativeness of the TS. Thus, under the same set size, the larger is the dimension of the space of features, the worse is the justification of the obtained statistical conclusions.

In the classification of a full image with application of the generated systems of 10 and 20 features (by the mentioned method based on the data of a TS fragment), we achieved the accuracy exceeding the classification accuracy in the system obtained by the regular decimation (by more than 7%) and on the basis of PCA, but somewhat ranking below compared to the system of features obtained with the MNF (Table 5).

5. CONCLUSIONS

Using numerical modeling, we demonstrated the possibility of determining reduced subsystems of features consisting of a small number of most informative hyperspectral channels (their position and width) based on their successive addition and removal.

During studies we experimentally showed that the registration and processing of data in 10 spectral channels chosen according to the proposed technique by a fragment of HS image allow obtaining the classification accuracy of a large-sized image in solving the problem of monitoring agricultural plants not worse than that with application of the principal component analysis. The labor intensity of the required computations decreases significantly. The choice of the spectral intervals and their number and position is performed at the stage of preparation to the solution of certain classes of problems and does not change during operation. Such approach leads to increasing efficiency of target problems due to the fact that it becomes possible to take into account the characteristic peculiarities of investigated objects and the observation conditions in each specific case sufficiently completely.

The results indicate that, for small sizes of training sets (experiments with a fragment of image), the application of more than 20 features does not lead to increase in the classification accuracy. In the case of extension of the TS (experiments with full image), the use of larger number of features allows somewhat increasing the classification accuracy, but causes a considerable increase in the hardware resources and computation time, which is not purposeful for most practical applications. This enables proceeding from application of expensive and complex hyperspectral apparatus, which is justified in space and near-Earth studies, to development and creation of small-sized devices of remote diagnostics with a small number of features specially selected for specific problems with the aim of their wide practical application.

One of the ways of further development is the rejection of sensors with fixed parameters and transition to adaptive systems of HS monitoring (for instance, with application of acoustooptical filters or electrically controlled Fabry–Perot interferometers). There are technical issues to be resolved in creating the new generation of systems: development and supply of new adaptive modes of apparatus functioning.

In these modes the information parameters are selected and adjusted for a specific problem (observation conditions) in real time. The number of adjustable information parameters includes the spatial and spectral resolutions and the bit rate of signal sampling. The realization of perspective modes of apparatus functioning allows significantly expanding the operation indicators of hyperspectral systems, thus improving the level of solving the observation and classification problems up to a qualitatively new grade.

REFERENCES

1. *Perspective Information Technologies of Earth's Remote Sensing*, Ed. by V. A. Soifer (Novaya Tekhnika, Samara, 2015).
2. V. G. Bondur, "Modern approaches to processing large hyperspectral and multispectral aerospace data flows," *Izv., Atmos. Ocean. Phys.* **50**, 840–852 (2014). <https://doi.org/10.1134/S0001433814090060>
3. V. V. Shipko, "Noise filtration in hyperspectral images," *Optoelectron., Instrum. Data Process.* **56**, 19–27 (2020). <https://doi.org/10.3103/S8756699020010033>
4. V. V. Kozoderov, T. V. Kondranin, and E. V. Dmitriev, "Recognition of natural and man-made objects in airborne hyperspectral images," *Izv., Atmos. Ocean. Phys.* **50**, 878–886 (2014). <https://doi.org/10.1134/S0001433814090126>
5. A. N. Vinogradov, V. V. Egorov, A. P. Kalinin, E. M. Melnikova, and A. I. Rodionov, "The set of hyperspectral sensors in the optical band," Preprint No. Pr-2176, IKI RAN (Space Research Institute, Russian Academy of Sciences, 2015).
6. N. Tack, A. Lambrechts, S. Soussan, and L. Haspelslagh, "A compact, high-speed, and low-cost hyperspectral imager," *Proc. SPIE* **8266**, 82660Q (2012). <https://doi.org/10.1117/12.908172>
7. B. Geelen, N. Tack, and A. Lambrechts, "A snapshot multispectral imager with integrated, tiled filters and optical duplication," *Proc. SPIE* **8613**, 861314 (2013). <https://doi.org/10.1117/12.2004072>
8. N. Gat, "Imaging spectroscopy using tunable filters: a review," *Proc. SPIE* **4056**, 50–64 (2000). <https://doi.org/10.1117/12.381686>
9. J. H. Correia, M. Bartek, and R. F. Wolffenbuttel, "High-selectivity single-chip spectrometer in silicon for operation in visible part of the spectrum," *IEEE Trans. Electron Devices* **47**, 553–559 (2000). <https://doi.org/10.1109/16.824727>
10. M. Jayapala, A. Lambrechts, N. Tack, B. Geelen, B. Masschelein, and P. Soussan, "Monolithic integration of flexible spectral filters with CMOS image sensors at wafer level for low cost hyperspectral imaging." http://www.imagesensors.org/Past%20Workshops/2013%20Workshop/2013%20Papers/07-02_053-jayapala.pdf. Cited May 15, 2020.
11. A. Shvedov, "Hyperspectral sensors of IMEC company. Solutions for high-quality spectral analysis." <https://www.npk-photonica.ru/giperspektralnye-sensory-kompanii-imec-pdf163406.pdf>. Cited May 15, 2020.
12. R. V. Skidanov, A. A. Morozov, A. P. Porfirev, and V. A. Blank, "An imaging spectrometer based on a discrete interference filter," *Comput. Opt.* **39**, 716–720 (2015). <https://doi.org/10.18287/0134-2452-2015-39-5-716-720>
13. S. M. Borzov, M. A. Guryanov, and O. I. Potaturkin, "Study of the classification efficiency of difficult-to-distinguish vegetation types using hyperspectral data," *Comput. Opt.* **43**, 464–473 (2019). <https://doi.org/10.18287/2412-6179-2019-43-3-464-473>
14. S. M. Borzov, A. O. Potaturkin, O. I. Potaturkin, and A. M. Fedotov, "Analysis of the efficiency of classification of hyperspectral satellite images of natural and man-made areas," *Optoelectron., Instrum. Data Process.* **52**, 1–10 (2016). <https://doi.org/10.3103/S8756699016010015>
15. S. M. Borzov and O. I. Potaturkin, "Efficiency of the spectral-spatial classification of hyperspectral imaging data," *Optoelectron., Instrum. Data Process.* **53**, 26–34 (2017). <https://doi.org/10.3103/S8756699017010058>
16. I. Fodor, "A survey of dimension reduction techniques," Technical Report No. UCRL-ID-148494 (Univ. California, Oakland, 2002).
17. I. A. Pestunov and P. V. Melnikov, "Principal component analysis and its modifications in hyperspectral image classification," *Interexpo Geo-Siberia* **4** (2), 45–50 (2015).
18. M. F. Baumgardner, L. L. Biehl, and D. A. Landgrebe, "220 band AVIRIS hyperspectral image data set: June 12, 1992 Indian Pine test site 3," (Purdue Univ. Research Repository, 2015). <https://doi.org/10.4231/R7RX991C>
19. T. Merill and D. Green, "On the effectiveness of receptors in recognition systems," *IEEE Trans. Inf. Theory* **JT**, 11–17 (1963). <https://doi.org/10.1109/TIT.1963.1057810>

20. P. Ghamisi and J. A. Benediktsson, "Feature selection based on hybridization of genetic algorithm and particle swarm optimization," *IEEE Geosci. Remote Sens. Lett.* **12**, 309–313 (2015). <https://doi.org/10.1109/LGRS.2014.2337320>
21. D. Beasley, D. R. Bull, and R. R. Martin, "An overview of genetic algorithms," *Univ. Comput.* **15** (2), 58–69 (1993).
22. Yu. L. Barabash, B. V. Varskii, V. T. Zinoviev, et al., *Automatic Image Recognition* (KVAIU, Kiev, 1964).
23. J. A. Richards, *Remote Sensing Digital Image Analysis* (Springer-Verlag, Berlin, 1999).
24. N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods* (Cambridge Univ. Press, Cambridge, 2000). <https://doi.org/10.1017/CBO9780511801389>
25. C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.* **2**, 27 (2011). <https://doi.org/10.1145/1961189.1961199>
26. N. G. Zagoruiko, V. N. Elkina, G. S. Lbov, and S. V. Emel'yanov, *Package of Applied Programs OTEKS* (Finansy i Statistika, Moscow, 1986).
27. V. N. Elkina, N. G. Zagoruiko, and V. S. Temirkaev, "Algorithms of directed taxonomic search of informative subsystems of features (DTSA)," *Vychislit. Syst. Sb. IM SO AN SSSR* **59**, 49–70 (1974).
28. *Remote Sensing: Qualitative Approach*, Ed. by A. S. Alekseev (Nedra, Moscow, 1983).
29. O. A. Dubrovskaya, T. A. Gurova, and I. A. Pestunov, "Disease detection on wheat plantings by data of hyperspectral monitoring," in *Proc. Int. Conf. Spatial Data Processing Monitoring Prob. Nat. and Human-Made Processes (SDM-2019), Novosibirsk, 2019*, pp. 497–500.
30. A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.* **26**, 65–74 (1988). <https://doi.org/10.1109/36.3001>

Translated by E. Oborin